

PARTIAL TRANSLATION OF JP 10 (1998)-124086 A

Publication Date: May 15, 1998

Title of the Invention: VOICE INTERACTIVE SYSTEM

Patent Application Number: 8-283461

Filing Date: October 25, 1996

Inventor: Kazuhiko TAKAHASHI

Applicant: MEIDENSHA CORP.

(Page 3, right column, lines 23-47)

[0018] Referring to Figure 1 again, regarding information recognized by a voice recognition device 15 and subjected to character conversion, meaning contents contained in a voice are extracted by a voice understanding device 16.

[0019] Unlike a conventional off-line knowledge base system, an expert system 17 has a function of inferring in real time while exchanging data with another process on the same computer under an on-line environment or another computer connected through a communication line.

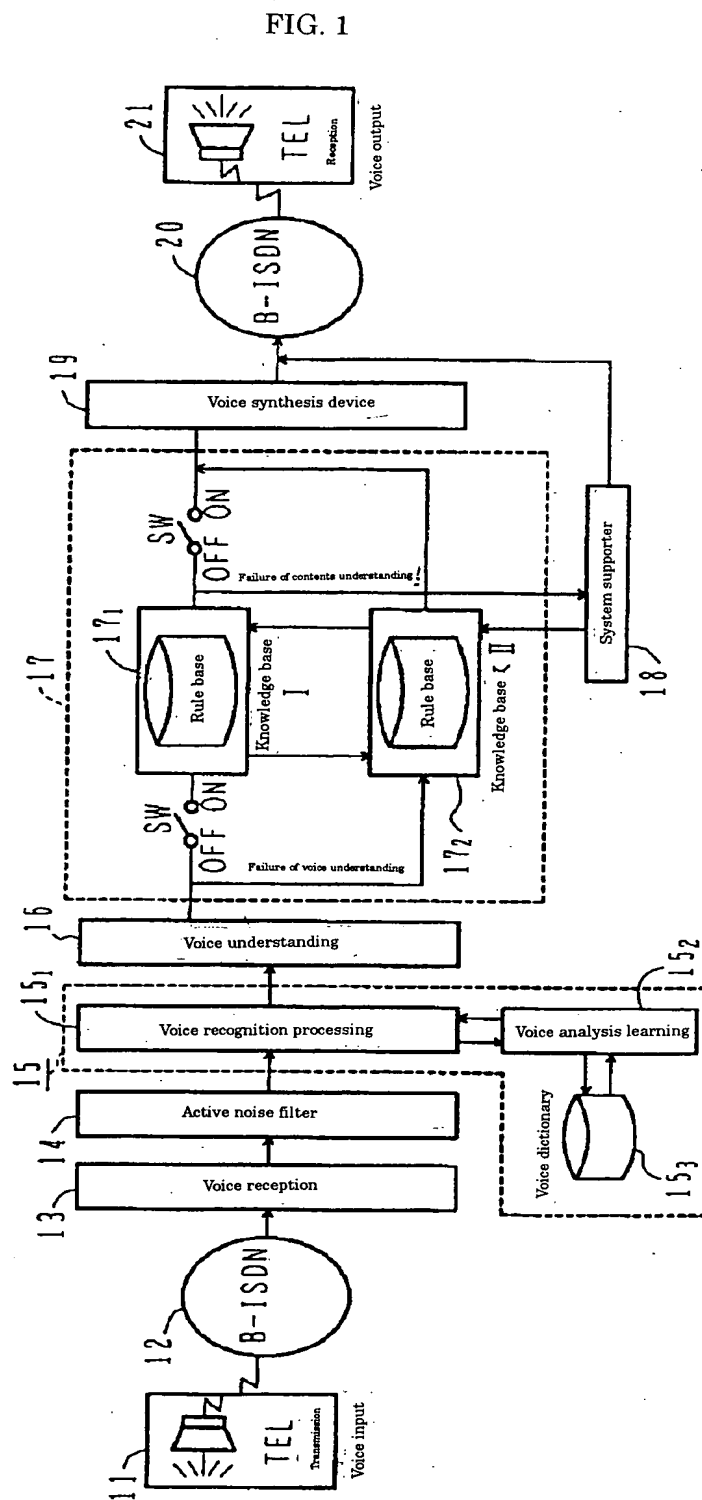
[0020] The expert system 17 includes a first knowledge base 17₁ and a second knowledge base 17₂ which are to be a knowledge base library. As shown in Figure 3, when a voice understanding device 16 fails in voice understanding, the expert system 17 infers, using a rule of the knowledge base 17₂, and obtains voice information of the expert system 17 notifying a user of failure of understanding. The expert system 17 has a learning function of increasing information stored in the knowledge base 17₂.

[0021] When the voice understanding device 16 understands voice contents, the expert system 17 infers, using a rule of the knowledge base 17₁, and obtains various pieces of voice information of the knowledge base 17₁ for responding to a user's request.

[0022] A system supporter 18 is used for conducting interactive processing mediated by a human, in the case where there is no voice information requested by a user in a knowledge base or a user's request cannot be

satisfied.

Block diagram of a system (Embodiment)



(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **10124086 A**(43) Date of publication of application: **15.05.98**

(51) Int. Cl.

G10L 3/00
G10L 3/00
G10L 3/00
G06F 3/16
// G06F 17/28

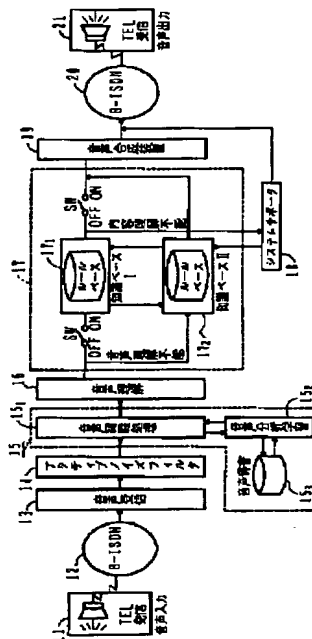
(21) Application number: **08283461**(71) Applicant: **MEIDENSHA CORP**(22) Date of filing: **25.10.96**(72) Inventor: **TAKAHASHI KAZUHIKO**(54) **VOICE INTERACTIVE SYSTEM**

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a system which can obtain the natural interaction between an unspecified speaker and a machine.

SOLUTION: A user's voice is received by a voice input device, a communication system 12 and a voice reception data 13, and an active noise filter 14 removes noise; and a voice recognition device 15 which uses unspecified-speaker word voice recognition technology performs speech recognition and a voice understanding device 16 extracts meaning contents from the voice. Then an expert system 17 having a knowledge base of information that users request and limited information obtains answer voice information for the users and when there is not the information, a person intervenes with a voice through a system supporter 18 to send speech information from a voice output device 21 to the users with a voice or voice signal synthesized by a voice synthesizing device 19.

COPYRIGHT: (C)1998,JPO



(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平 10 - 124086

(43) 公開日 平成10年(1998)5月15日

(51) Int. Cl. ⁶	識別記号	F I
G 1 0 L 3/00	5 7 1	G 1 0 L 3/00 5 7 1 H
	5 6 1	Q
G 0 6 F 3/16	3 1 0	5 6 1 G
// G 0 6 F 17/28		G 0 6 F 3/16 3 1 0 A
		15/38 V
審査請求 未請求 請求項の数 1	OL	(全 10 頁)

(21) 出願番号 特願平8-283461

(22) 出願日 平成8年(1996)10月25日

(71) 出願人 000006105

株式会社明電舎

東京都品川区大崎2丁目1番17号

(72) 発明者 高橋 和彦

東京都品川区大崎2丁目1番17号 株式会社

明電舎内

(74) 代理人 弁理士 志賀 富士弥 (外1名)

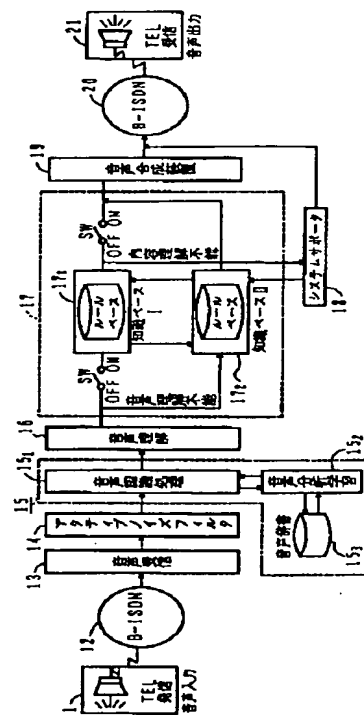
(54) 【発明の名称】 音声対話システム

(57) 【要約】

【課題】 音声合成装置や音声認識・理解システム等を単に組み合わせた音声対話システムでは、不特定の人と人があたかも自然に会話をしているようなヒューマンインタフェースを実現できない。

【解決手段】 利用者の音声を音声入力装置 11 と通信システム 12 と音声受信装置 13 で受信し、アクティブノイズフィルタ 14 でノイズを除去し、不特定話者単語音声認識技術を用いた音声認識装置 15 により音声認識し、音声理解装置 16 により音声に含まれている意味内容を抽出し、利用者が要求している情報や限定情報の知識ベースを持つエキスパートシステム 17 により利用者に対する応答音声情報を得、この情報が無いときにシステムサポート 18 により人が音声で媒介し、音声情報を音声合成装置 19 により合成した音声又は音声信号を通信システム 20 を介して音声出力装置 21 から利用者に音声で応答する。

システムのブロック図 (実施形態)



【特許請求の範囲】

【請求項1】 利用者の音声を通信システムを介して受信し、雑音を除去した音声信号を得る音声信号受信手段と、

前記音声信号のA/D変換データを周波数スペクトルによって大まかに分析し、特徴抽出によって音声信号の時系列の特徴パラメータを音素単位に抽出し、各音素単位の認識に続く単語の認識、及び構文認識と意味認識により音声信号に対応した文字データを得る音声認識装置と、

前記文字データの意味内容を抽出する音声理解装置と、前記音声理解装置が文字データの意味内容の理解に失敗したときに知識ベースを使った推論で音声不理解の音声情報をオンラインで得、意味内容を理解できたときに知識ベースを使った推論で利用者の要求に答えるための音声情報をオンラインで得るエキスパートシステムと、利用者の要求する音声情報が前記知識ベースに無い場合や対応できない場合に人が媒介して対話処理を行うためのシステムサポートと、

前記エキスパートシステムからの音声情報を音声信号として合成する音声合成装置と、

前記音声合成装置からの音声信号又はシステムサポートからの音声信号を通信システムを介して利用者側に伝送する音声信号送信手段と、を備えたことを特徴とする音声対話システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、不特定話者とコンピュータの間で音声により対話を行う音声対話システムに関する。

【0002】

【従来の技術】 従来、人と機械との間の対話には、テープレコーダ等に予め録音されたさまざまな情報を電話で聞き取ったり、ファクリミリを利用した情報収集を行うときに、アナウンスに従って人が機械操作を行っている。この場合、機械側が発生できる音声情報が限られるし、人は機械操作で応答するため、対話内容及び応答時間が大きく制限される。

【0003】 これら制限を無くそうとするものとして、機械側が発生できる音声情報を任意とする規則合成方式による音声合成装置がある。また、人が音声で応答し、これを機械側が認識する音声認識・理解システム、さらには画像情報と音声情報を取り扱うマルチメディア向け音声インタフェースがある。

【0004】 図7は、日本語規則合成方式の音声合成処理ブロック図を示す。日本語解析処理部1は、パソコンやワープロ、ワークステーション等から入力された漢字カナ混じり文に対して日本語辞書2を参照した構文解析、意味解析等により読み・単語・文節境界等を解析する。この解析結果から、抑揚制御部3による音韻系列の

アクセントとイントネーション及び継続時間長制御部4による音韻の継続時間の各パラメータを決定し、これらを音響管モデルのパラメータとして音声合成部5により音声辞書6を参照した音声合成を行い、拡声器7から合成音声を得る。

【0005】 このような音声合成装置を対話システムに利用するには、回線外部からの接続応答、回線外部への接続要求、電話機のプッシュボタン信号認識機能などの回線制御機構を持つ装置に音声合成装置を設け、その音声合成部5のモジュラー端子から直接に電話回線へ出力する。この装置は、コンピュータに接続するのみで電話対応システムが構築でき、専用のラックを用いて最大32回線までの同時応答や一斉通報ができる。

【0006】 次に、音声認識装置は、入力される音声の音響分析と音韻認識と単語認識及び言語処理などにより、単語音声認識、連続単語音声認識や音声理解を行う。例えば、単語抽出型音声認識システムは、図8に示すように、ファジィ推論で必要な単語を大つかみに抽出し、必要とする単語は予め登録しておき、後から学習機能によって次第に追加して行くシステムであり、ソフトウェアでも作動させることが可能であるため、コンピュータと人の対話における音声認識手段として利用することができる。

【0007】 マルチメディア向け音声インタフェースとしては、図9に示すように、動画像処理部と音声応答を用いたマルチメディア技術の実現の一環として、VMEバス対応高速ニューロコンピュータ上で開発した高精度な不特定話者単語音声認識技術を利用し、マルチタスク対応の音声入力インタフェースをワークステーションのXウインドウ上に構築したものがある。

【0008】 このシステムの特徴としては、(a)フォーカスウインドウに対する音声入力が可能、(b)タスクに対応した音声辞書切換が可能、(c)アプリケーションに対してマウスと等価な動作が可能などがあげられる。

【0009】

【発明が解決しようとする課題】 人と機械との対話手段として、従来から音声合成装置や音声認識・理解システム、さらにはマルチメディア向け音声インタフェースがあるが、これらを単に組み合わせて音声対話システムを構築するのでは、不特定のひとと人があたかも自然に会話をしているようなヒューマンインタフェースを実現できない。

【0010】 本発明の目的は、不特定話者と機械の間で自然な対話を得ることができるシステムを提供することにある。

【0011】

【課題を解決するための手段】 本発明は、あたかも人と人が自然に会話をしているようなヒューマンインタフェースを実現するために、不特定話者に対応可能な技術と

してVMEバス対応高速ニューロコンピュータ上で開発した高精度な不特定話者単語音声認識技術を用い、かつ利用者が要求している情報や限定情報を拡張するためにルールによって対応可能とする知識ベース及び学習機能を持ったエキスパートシステム、さらに利用者の要求を返答するためより自然で人間の肉声に近い音声出力を得る音声合成システムを統合し、かつワークステーション上で動作可能としたヒューマンインタフェースを実現するもので、以下の構成、利用者の音声を通信用システムを介して受信し、雑音を除去した音声信号を得る音声信号受信手段と、前記音声信号のA/D変換データを周波数スペクトルによって大まかに分析し、特徴抽出によって音声信号の時系列の特徴パラメータを音素単位に抽出し、各音素単位の認識に続く単語の認識、及び構文認識と意味認識により音声信号に対応した文字データを得る音声認識装置と、前記文字データの意味内容を抽出する音声理解装置と、前記音声理解装置が文字データの意味内容の理解に失敗したときに知識ベースを使った推論で音声不理解の音声情報をオンラインで得、意味内容を理解できたときに知識ベースを使った推論で利用者の要求に答えるための音声情報をオンラインで得るエキスパートシステムと、利用者の要求する音声情報が前記知識ベースに無い場合や対応できない場合に人が媒介して対話処理を行うためのシステムサポータと、前記エキスパートシステムからの音声情報を音声信号として合成する音声合成装置と、前記音声合成装置からの音声信号又はシステムサポータからの音声信号を通信用システムを介して利用者側に伝送する音声信号送信手段と、を備えたことを特徴とする。

【0012】

【発明の実施の形態】図1は、本発明の実施形態を示すシステムブロック構成図であり、図2～図4に各部の処理過程を同じ符号で対応付けて示す。以下、各部の構成と処理を詳細に説明する。

【0013】音声入力装置11は、電話機など受話器のマイクロフォンから利用者が音声を入力するためのものである。通信用システム12は、例えば、デジタル回線で高速かつ広帯域性がある2B+DのB-ISDN及びネットワークを制御するATM交換システムによって通信制御を行い、音声入力装置11からの音声信号を音声受信装置13に伝送する。アクティブノイズフィルタ14は、音声受信装置13が受信した音声信号から雑音を除去する。

【0014】音声認識装置15は、受信した音声进行認識及び理解するため、不特定話者に対応可能な技術としてVMEバス対応高速ニューロコンピュータ上で開発した高精度な不特定話者単語音声認識技術のニューロコンピュータシステムに構成され、音声認識処理部15₁と音声分析学習部15₂及び音声辞書15₃を備える。

【0015】この装置15では、図2及び図3に音声認

識ライブラリを使用した処理過程で示すように、入力された信号を音声理解処理、A/D変換処理した後、音声分析により音声入力信号を周波数スペクトルによって大まかに分析し、特徴抽出によって音声分析結果から音声信号の特徴パラメータを時系列に抽出・変換し、セグメンテーション処理により音声単位へのセグメント化を行う。

【0016】次いで、音素認識により音素標準パターンとの比較で音素認識を行うことで音素系列を得、単語照合と単語認識により音素系列について知識ベースに持つ単語標準パターンとの照合で単語を認識する。この認識に標準パターンが存在しなければ学習処理（認識・登録）を行って単語知識を知識ベースに追加する。

【0017】次いで、認識した単語について、構文照合と構文認識により知識ベースの構文パターンとの照合で構文的に誤りがあるか否かを解析し、誤りがあれば再検証することで構文認識を行う。さらに、意味解析と意味認識により認識された単語構文について意味的に妥当であるかを調べ、妥当な結果が得られるまで構文解析処理と意味解析処理を繰り返す。これら解析には文字に変換可能か否かも含め、文字変換処理により変換可能である文字を漢字やカナに変換する。

【0018】図1に戻って、音声認識装置15により認識されかつ文字変換を終えた情報について、音声理解装置16により音声に含まれている意味内容を抽出する。

【0019】エキスパートシステム17は、従来のオフライン型の知識ベースシステムと異なり、オンライン環境下で同一コンピュータ上の他のプロセスや通信回線で結合された他のコンピュータとデータをやり取りしながら実時間で推論を行う機能を持つ。

【0020】このエキスパートシステム17は、知識ベースライブラリになる第1の知識ベース17₁と第2の知識ベース17₂を備え、第3図に示すように、音声理解装置16による音声理解が失敗したときには知識ベース17₂のルールを使った推論を行い、この知識ベース17₂が持つ音声不理解であることを利用者に知らせる内容の音声情報を得る。なお、エキスパートシステム17は、学習機能を有して知識ベース17₂への保存情報を増大させていく。

【0021】エキスパートシステム17は、音声理解装置16が音声内容を理解できたときには、知識ベース17₁のルールを使った推論を行い、この知識ベース17₁が持つ利用者の要求に答えるためのさまざまな音声情報を得る。

【0022】システムサポータ18は、利用者の要求する音声情報が知識ベースに無い場合や対応できない場合に人が媒介して対話処理を行うのに使用する。

【0023】音声合成装置19は、エキスパートシステム17からの音声情報を音声信号として合成する。この装置19は、従来の音声合成装置と同等のものになり、

より自然で人の肉声に近い音声出力を得る。

【0024】通信システム20は、図4に示すように、音声合成装置19により合成した音声信号、又はシステムサポータ18からの生の音声をB-ISDNとATM交換システム等によって利用者側に伝送し、音声出力装置21により音声を再生する。

【0025】以上までのシステムブロック構成に必要なインタフェースは、図5に示すようになる。また、システム構成は図6に示すようになる。

【0026】以上までの構成により、本実施形態では、利用者とは機械との間であたかも人と人が自然に会話をしているようなヒューマンインタフェースを実現する。

【0027】

【発明の効果】以上のとおり、本発明によれば、不特定話者に対応可能な不特定話者単語音声認識技術を用い、かつ利用者が要求している情報や限定情報を拡張するためにルールによって対応可能とする知識ベース及び学習機能を持ったエキスパートシステム、さらに利用者の要求を返答するためより自然で人間の肉声に近い音声出力を得る音声合成システムを統合したため、人と機械の対話に不特定の人と人があたかも自然に会話をしているようなヒューマンインタフェースを実現することができる。

【図面の簡単な説明】

【図1】本発明の実施形態を示す音声対話システムのブロック構成図。

【図2】図1における処理過程（その1）。

【図3】図1における処理過程（その2）。

【図4】図1における処理過程（その3）。

【図5】実施形態におけるインタフェース構成例。

【図6】実施形態におけるシステム構成例。

【図7】ユニットタイプの音声合成処理ブロック図。

【図8】単語音声認識のスペクトラム及び抽出例。

【図9】音声認識インタフェース例。

【符号の説明】

11…音声入力装置

12、20…通信システム

13…音声受信装置

14…アクティブノイズフィルタ

15…音声認識装置

16…音声理解装置

17…エキスパートシステム

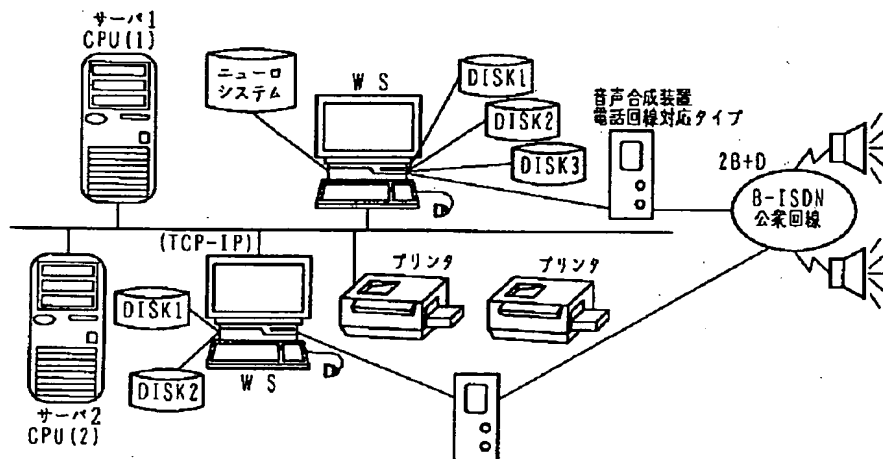
18…システムサポータ

19…音声合成装置

21…音声出力装置

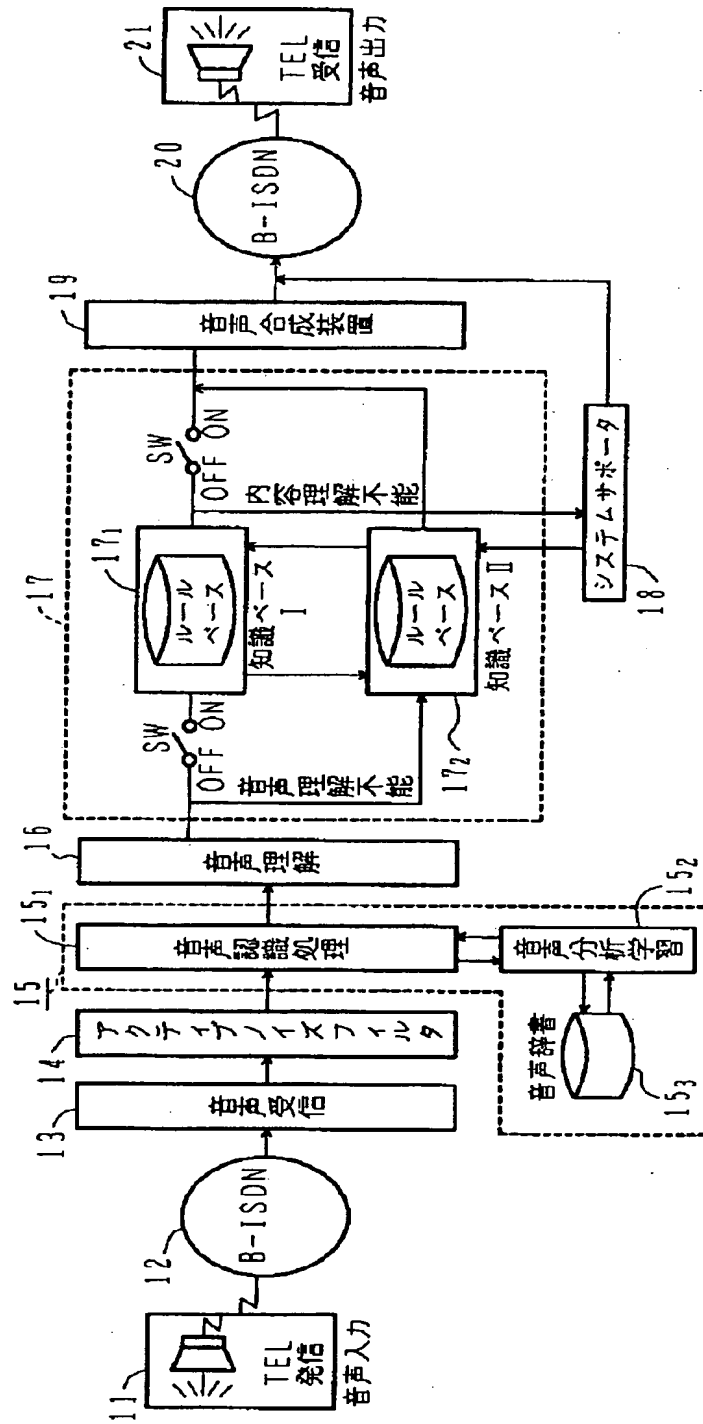
【図6】

システム構成例



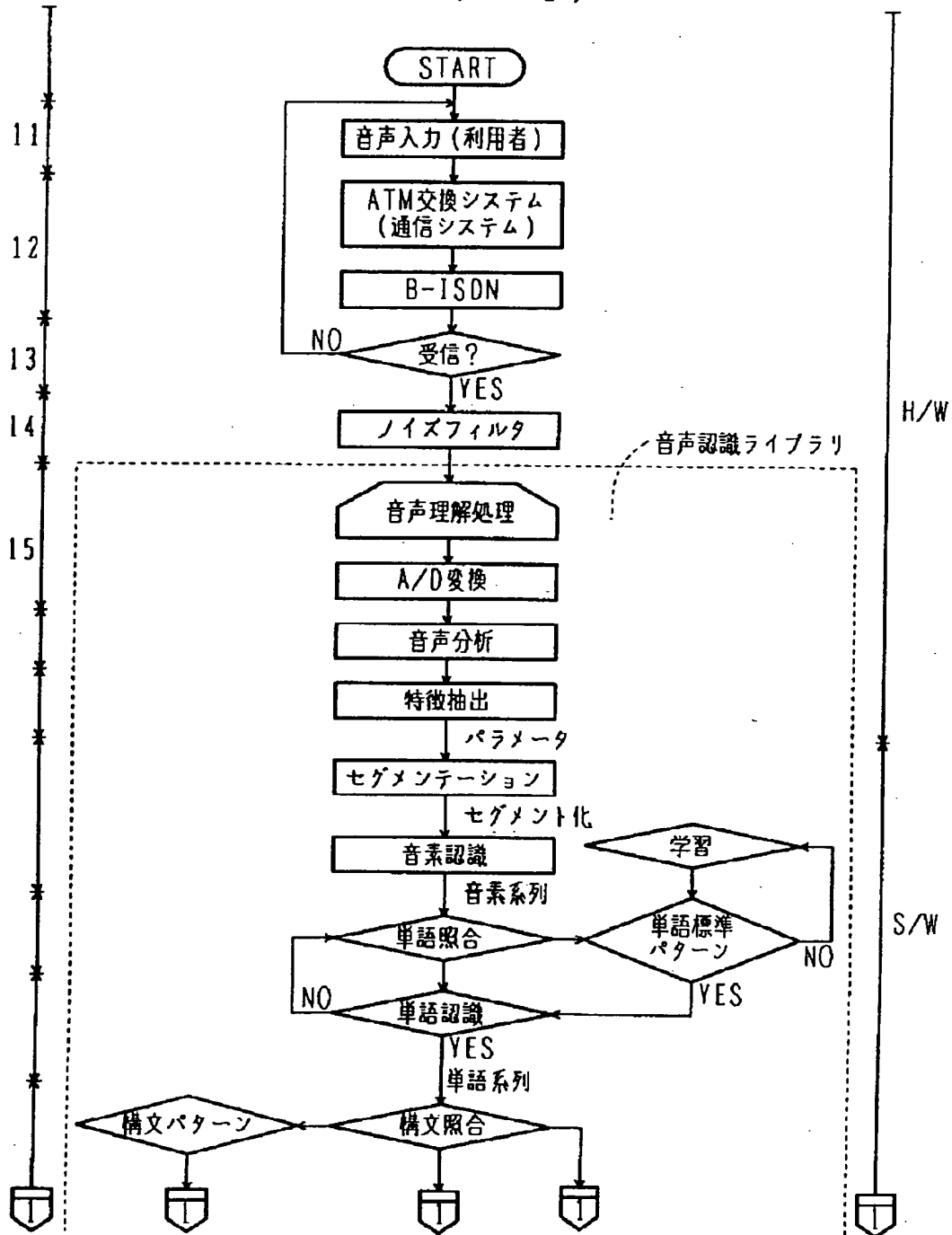
【図1】

システムのブロック図（実施形態）



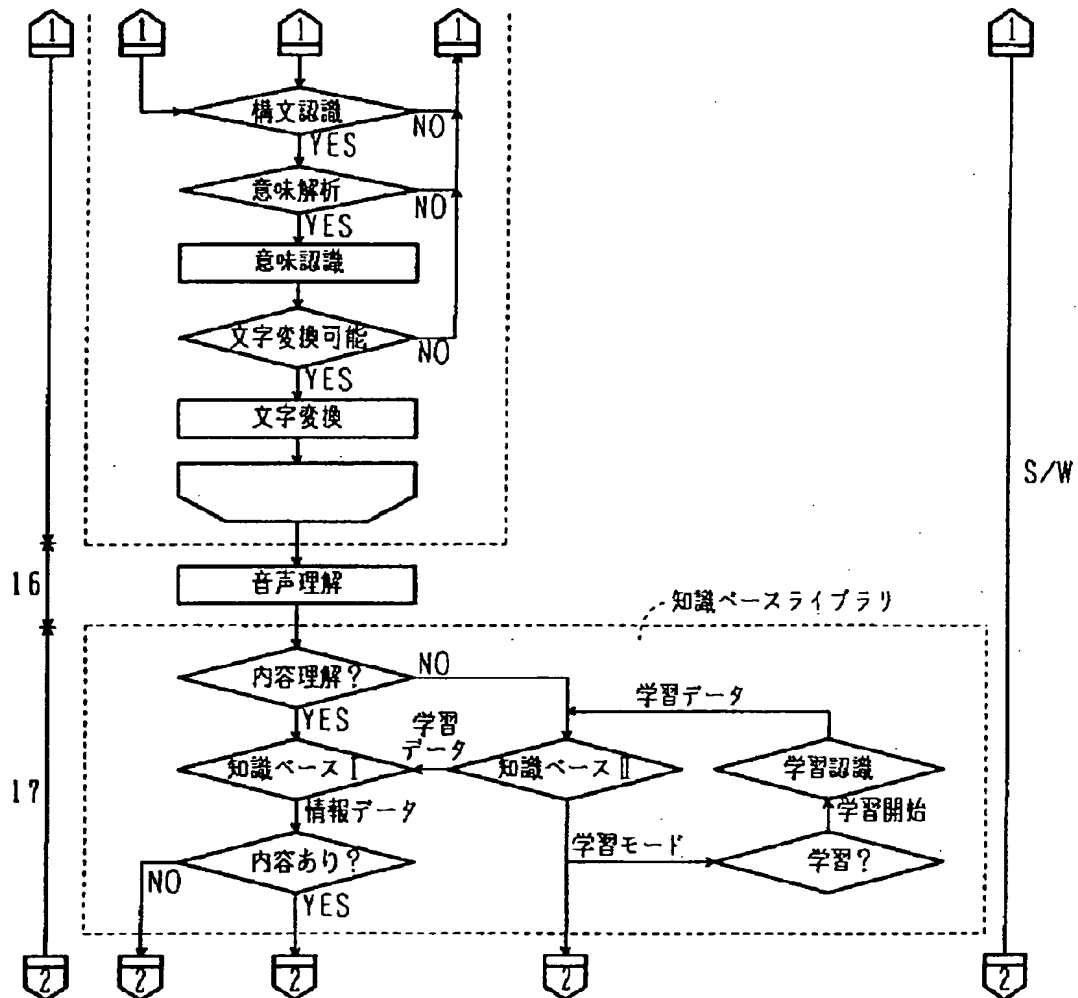
【図2】

処理過程（その1）



【図3】

処理過程（その2）



【図8】

単語音声認識のスペクトラム及び抽出例

えーと、大月まで案内地図一枚をお願いします。

録音・音声および
必要単語の分析



単語抽出

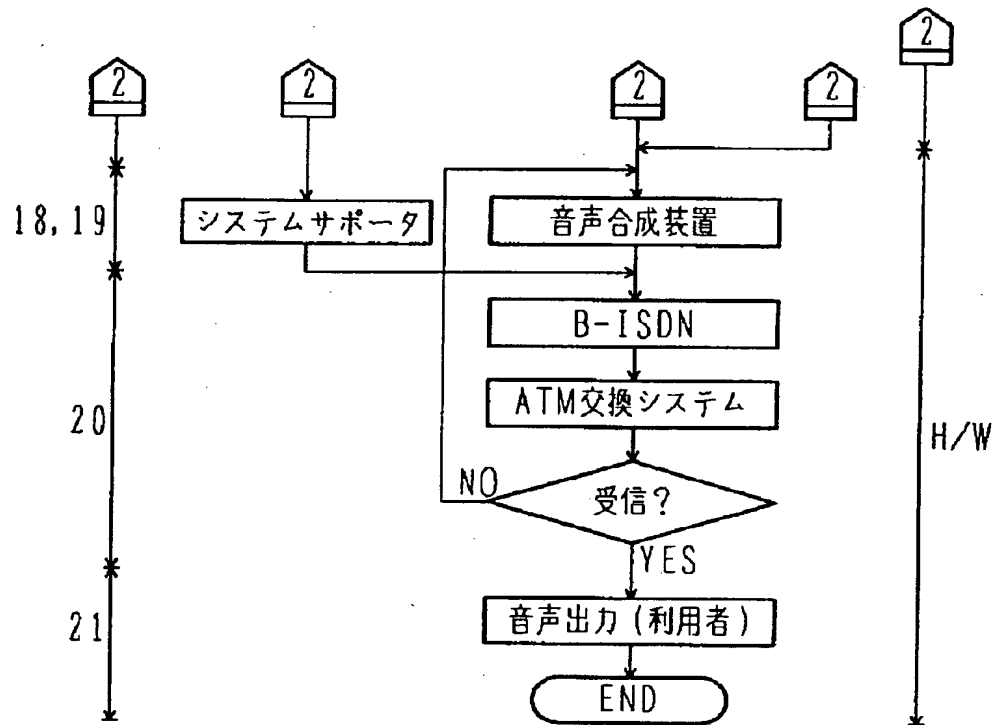
|おおつき| |あんない| |ちず| |いちまい|

単語の決定

大月 案内 地図 一枚

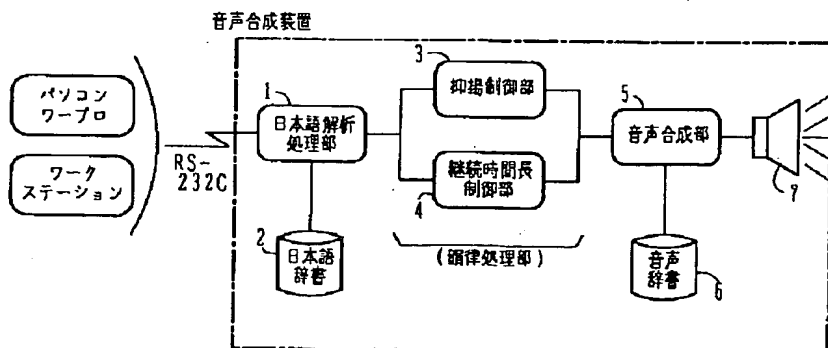
【図4】

処理過程（その3）

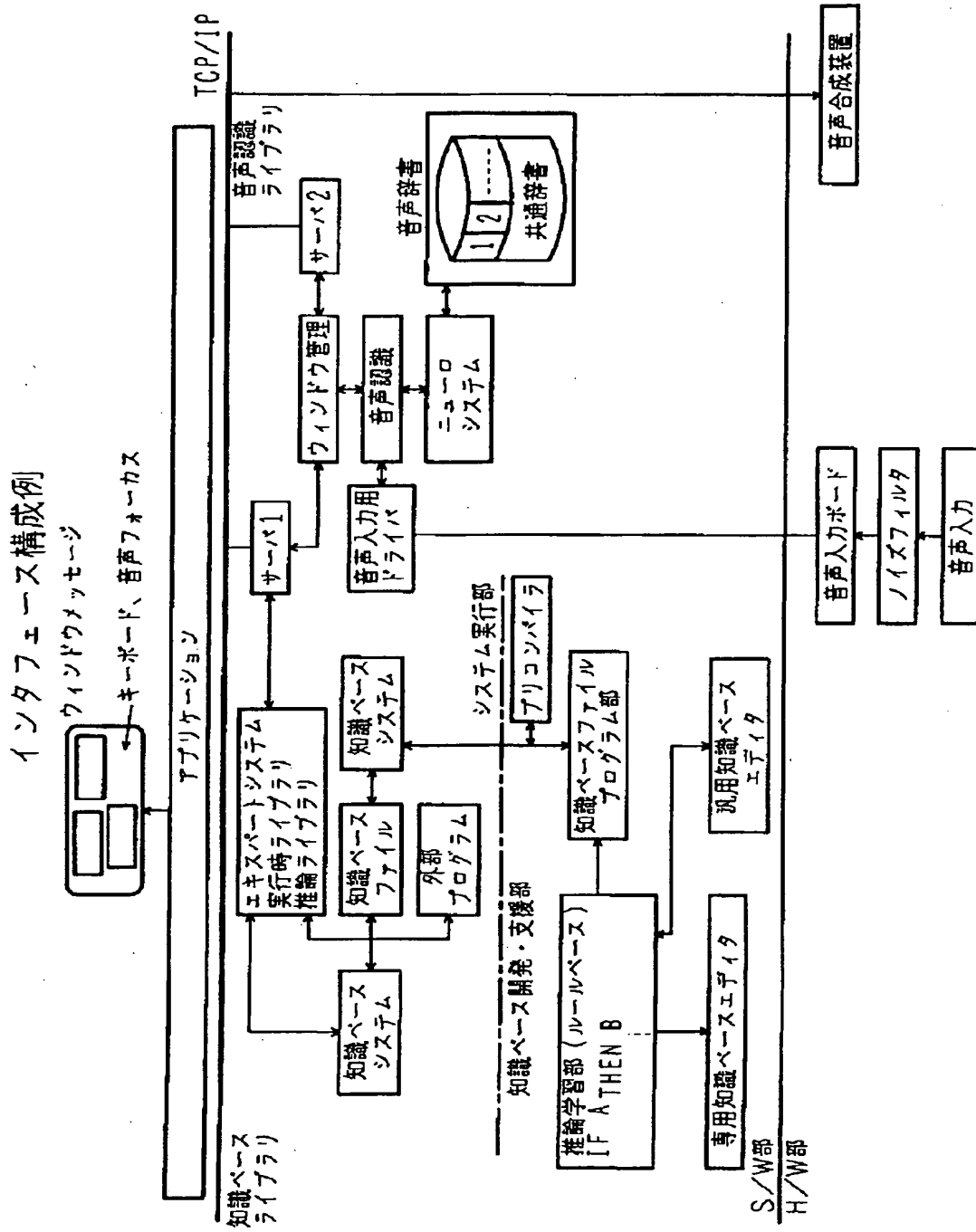


【図7】

ユニットタイプの音声合成処理ブロック図



【図5】



【図9】

音声認識インタフェース例

